

## The data and the examples: Comprehensiveness, accuracy, and sensitivity

Marianne Mithun

*University of California, Santa Barbara*

Good grammars are read by diverse audiences with a wide variety of interests. One might not write a reference grammar in exactly the same way for all potential users, but particularly in the case of under-documented and endangered languages, it is likely that whatever is produced now will be consulted for answers to questions beyond those originally anticipated. A good grammar can provide more than descriptions of patterns the grammarian has noted at the time of writing; the examples it contains can provide a basis for future discoveries and new uses. It thus makes sense to consider the types of data that might best meet the needs of current and future readers, some of which we cannot even imagine at present. For some purposes, sensitive, typologically-informed elicitation is necessary, while for others, material drawn from unscripted connected speech is crucial. Here the potential contributions of examples of each type are considered for descriptions of phonetics, phonology, morphology, syntax, discourse, prosody, language change, and language contact.

Writing a grammar can be one of the most demanding projects a linguist undertakes, but also one of the most exhilarating. On the one hand, it requires a broad mastery of general linguistics, a deep understanding of the language to be described, and dedication to a potentially monumental task. On the other, it can be immensely satisfying: an opportunity to see beautiful systems in all their richness and complexity, to watch their interactions, and to appreciate the language as a whole. Especially for those writing a grammar for the first time, there is much to think about before plunging in: the intended audience, the topics to be covered, the organization, the style, and more. The focus here will be on just one aspect of the enterprise: the data that form the heart of the work and the selection and presentation of the material in examples. If the data come from direct collaboration with speakers, it is useful to reflect on the kinds of examples that will form the basis of the grammar during both the documentation process and the grammar-writing process.<sup>1</sup>

Readers have traditionally approached reference grammars with certain expectations, anticipating basic descriptions of the sound system, word formation patterns, and sentence structures. Such information is still fundamental to a good grammar. But as the field of linguistics progresses and the community of grammar readers widens, ideas are evolving about the kinds of information that can move the field ahead and serve the needs of more users. Theoretical and technological advances are stimulating us to ask new questions and

1 I would like to express thanks to Toshihide Nakayama and Hideo Sawada, who organized the International Symposium on Grammar Writing as part of the Linguistic Dynamics Science Project at the ILCAA, Tokyo University of Foreign Studies. The symposium stimulated us all to think more deeply about the process of grammar writing and prompted useful discussion. I am especially grateful to the Mohawk speakers who have contributed their insights into the issues discussed here, in particular Charlotte Bush, Josephine Horne, Kaia'titáhkhe' Jacobs, and Watshenni:ne' Sawyer.

providing new tools for answering them. At the same time, the looming threat of language loss is inspiring more and more communities to mount language programs aimed at the documentation, preservation, and especially revitalization of their heritage languages. Distinctions between academic and community scholars are blurring: dedicated scholars coming from both directions are bringing increasing expertise and sophistication to their work. All are interested in clear, rich descriptions of the language, though sometimes their priorities differ. All of these developments mean that more kinds of readers are approaching grammars with a wider range of hopes.

Good grammars are read by a variety of linguists, including those specifically interested in phonology, morphology, syntax, discourse, typology, language change, and language contact, as well as relations among language, culture, and thought, and more. Reference grammars are increasingly consulted by those undertaking language revitalization projects, as a basis for planning language classes, preparing classroom materials, developing larger scale curricula, and compiling pedagogical grammars. They may be picked up by other community members simply interested in their heritage, by cultural anthropologists, by ethnohistorians, and others. One might not write a reference grammar in exactly the same way for all of these audiences: a grammar aimed at syntacticians might differ from one designed for language teachers. One written for typologists might differ from one aimed at interested community members. But particularly in the case of underdocumented and endangered languages, it is likely that whatever is produced will be consulted by a wider audience than originally anticipated. It is thus useful to try to imagine, when assembling the data that will form the basis of the grammar and choosing the examples that will ultimately appear in it, the kinds of information that might best meet the varied and evolving needs of current and future users.

There was a period in linguistics when sparseness was highly valued: students were sometimes told that no more examples should be included in a discussion than the bare minimum necessary to justify particular statements. Now, with more widespread recognition that the issues of concern to linguists are constantly evolving, and the acceleration of language loss in the world, the importance of rich data is becoming increasingly recognized. The examples in a grammar should provide evidence for the generalizations made, but they can also serve as a basis for further discoveries. The number and complexity of examples that can and should be included in a grammar necessarily vary from one situation to the next. For some non-linguists, short descriptions with small numbers of simple examples may be more accessible. Sometimes there are practical limitations on the length of a printed book. But if the grammar is to be the only description of the language, particularly if the language is threatened, other issues may be in play. Quantity is crucial for revitalization projects: language teachers and learners need more than a few forms to understand and absorb a pattern. Because the audience for a good grammar will be diverse, examples will be consulted for a variety of purposes, often beyond those envisioned by the author. Every example must thus be accurate on all levels: phonological, morphological, syntactic, lexical, and pragmatic.

Of course different languages show complexity in different areas of structure and use. Accordingly, good grammars will vary not just in the complexity of their examples, but also in the distribution of this complexity. Some points to consider when assembling data and examples for a grammar are discussed in the following sections. Not all are equally

relevant for all languages, all situations, or all times: languages are disappearing, the circumstances in which they are spoken are evolving, and technological possibilities are expanding. Many of the points discussed here are illustrated with examples from Mohawk, an Iroquoian language from northeastern North America, but for the most part, the details of the examples are less important than the principles they were chosen to exemplify.

**1. BASIC DOCUMENTATION.** If the grammar is to be based on material assembled by the grammarian in collaboration with speakers, it is useful to begin thinking about the kind of data that will provide a good foundation at the outset of the documentation work. With accelerating progress in technologies for audio/video recording and data management has come increased attention to procedures for language documentation. Himmelmann characterizes the field of language documentation as ‘concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties’ (2006:v). Discussions of issues involved in documentation can be found in Chelliah (2001), Mithun (2001, 2007), and Chelliah and de Reuse (2011), as well as the collections in Gippert, Himmelmann & Mosel (2006), Grenoble & Furbee (2010), the Language Documentation & Description series edited by Peter Austin and published by the Hans Rausing Endangered Languages program, and especially in the online journal *Language Documentation & Conservation* (<http://nflrc.hawaii.edu/ldc/>). These works cover such topics as kinds of data that should go into the record, recommendations for recording equipment and practices, formulas for metadata, formats for portability, preservation, and dissemination, ethics, and models of collaboration between speech communities and outside scholars. These collections also provide useful lists of web resources, links to software for transcription and annotation, metadata and corpus management, fonts, encodings and keyboard assignments, and speech analysis.

Not all of the data collected during a documentation project will figure equally in a grammar. Neither elicited nor spontaneous data are sufficient on their own, and the optimal balance between the two is an ongoing matter of discussion. Elicitation can provide orderly paradigms and parallel forms for comparisons. But also important is as large a corpus as possible of unscripted connected speech in a range of genres, both monologue of various types and especially conversation. Different genres can provide evidence of different aspects of the language. Furthermore, spontaneous speech is more likely to be idiomatic, providing examples not just of basic grammatical structures, but of grammatical constructions along with the kinds of lexical substance that is conventionally attached to them.

**2. BASIC PRESENTATION OF EXAMPLES.** A very useful discussion of the selection and presentation of examples for grammars is in Weber (2007). Weber points out that examples must illustrate the claim being made, they must be accurate, they should illustrate a range of uses, they should be structurally diverse, and, where possible, they should be culturally interesting. One should be certain that they are free of gender bias, project a good image of the speakers and their culture, and do not embarrass particular individuals or groups. Weber provides some ordering principles, suggesting that examples that best illustrate the claim being made be ordered first, that simple examples appear before more complex ones, that more typical ones precede more unusual (marked) ones, and that ambiguous cases (those

that could be interpreted either in such a way as to support the claim or in some other way) be ordered last.

Weber further points out that examples should be framed in the grammar: their relevance to the point being made should be explained in the prose immediately before or after them. Claims and examples should be integrated, in order to avoid lengthy descriptions followed by lengthy series of examples.

Choices about the layout of examples and the amount of information to include with them will vary with the nature of the language, the situation, and the points being made. Particularly for languages with morphological complexity, examples are typically presented in a multi-line format. Some common conventions for formats and abbreviations for grammatical terminology are laid out in the Leipzig Glossing Rules, assembled by linguists at the Max Planck Institute for Evolutionary Anthropology and available at <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>.

An example from Mohawk, which can show relatively elaborate morphology, is in (1).

(1) Mohawk example: Watshenní:ne' Sawyer, speaker p.c., Bridge 2.43<sup>2</sup>

<i>Tsi nnió:re'</i>	<i>tsi kahnekí:io</i>
tsi ni-io-r-e'	tsi ka-hnek-iio
so PRT-N.PAT-be.far-STATIVE	so N.AGT-liquid-be.good-STATIVE
so it is that far	so it is liquid good
'The water was so good that	
 <i>enwá:ton'</i>	 <i>nenhshnekì:ra'.</i>
en-w-aton-'	ne=en-hs-hnek-ihra-'
FUT-N.AGT-be.possible-PFV	the=FUT-2SG.AGT-liquid-consume-PFV
it will be possible	the you will liquid consume it
you could drink it.'	

Here the top line presents the utterance as spoken, in the standard community orthography. Some publishers require that this line be in italics, while others prefer a basic roman font. Sometimes boldface is preferred, especially by communities who want to highlight the importance of the language being described.

- 2 Abbreviations for glosses follow the Leipzig Glossing Rules (<http://www.eva.mpg.de/lingua/resources/>). Additional abbreviations are AGT GRAMMATICAL AGENT, CONTR CONTRASTIVE, I INDEFINITE OR GENERIC GENDER, N NEUTER GENDER, PAT GRAMMATICAL PATIENT, PRT PARTITIVE, Z ZOIC GENDER. Unless otherwise noted, Mohawk examples are given in the standard practical orthography. The phonetic values of the symbols are essentially like those of the IPA, with the following exceptions. Orthographic <i> is a palatal glide [j] before vowels, but the usual high front unrounded vowel [i] elsewhere. The digraph <en> represents a nasalized caret [ʌ], and the digraph <on> a nasalized high back rounded vowel [ʊ]. The colon <:> indicates vowel length. Acute accents <ó> indicate stress and rising tone, and grave accents <ò> stress and falling tone. The apostrophe <'> represents glottal stop [ʔ]. Stops <t> and <k> are automatically voiced before other voiced segments. The sequence <ti> before a vowel is pronounced as an affricate [dʒ]. Further abbreviations are EP epenthetic vowel; LK linker (the vowel -a- inserted between noun and verb stems in compounds under certain phonological conditions, and between verb stems and derivational suffixes).

The next line is a parsed line, showing the internal morphological structure of each word. Here boundaries between morphemes are indicated with hyphens: en-w-aton-'. Boundaries between clitics and their hosts are shown by an equals sign: ne=en . . . If the language shows extensive phonological processes, the grammarian can decide how abstract the representations of the morphemes should be. The particular Mohawk words in (1) do not show many phonological processes. When the article *ne* occurs before a word beginning in a vowel, it often cliticizes and is reduced to *n=*. In (1) it is represented as spoken on the top line (*n*), but shown in its full form on the parsed line (*ne=*). Stress placement and vowel length are not properties of individual Mohawk morphemes, but rather of full words. Neither is represented in the basic forms of the morphemes on the second line.

The third line provides glosses, the meaning or function of each morpheme. Lexical glosses are given in roman type (be.good), and grammatical terms are given in small caps (STATIVE). When a single morpheme in the second line corresponds to a multi-word gloss in the third line, the words in the gloss are linked by periods: the gloss for the Mohawk *-iio*, for example, is given here as be.good.STATIVE. No period separates person and number, as in 2SG, the gloss for the second person singular pronominal prefix *-hs-* 'you'.

Because Mohawk morphology is sometimes complex, it can be useful to readers to have a word-by-word translation as well, as on the fourth line in (1).

Finally, the last line provides a free, idiomatic translation, usually surrounded by single quotation marks. It is important to keep all associated lines together on a page. Particularly when the order of information in the target language contrasts strongly with that in the free translation, it may be easier to present the full free translation as a separate unit at the end of the example.

Where publishers permit, it can be useful to use a smaller font for the interlinear analysis lines, like the second, third, and fourth lines in (1). Some readers are less interested in the analysis and find these lines distracting. In some electronic formats it is possible for readers to see only as much as they wish on a computer screen at one time. They may even be able to hover over certain elements to bring up further information, such as the morphological analysis of words, additional forms and/or meanings of a morpheme, the discourse context surrounding the example, or even sound. And possibilities are increasing all the time.

The Mohawk sentence in (1) is presented such that each group of lines represents a separate prosodic phrase or intonation unit, that is, it is arranged by prosodic rather than syntactic structure. Such an arrangement can be useful, displaying structure that would be lost if examples were broken into lines on a purely syntactic basis.

The layout in (1) is of course not appropriate for every purpose or every language. In a section of a grammar that lists distinctive sounds, single words and glosses are usually sufficient. In a paradigm, simple lists of related words followed by their meanings are usually most effective. In a discussion of lexical categories, on the other hand, an additional line might be useful that identifies parts of speech. If a language is usually written with a non-roman orthography, such as the Cherokee syllabary or Chinese characters, an additional line of transcription might be important. A language might have such simple phonological structure that a separate parsed line is not necessary: the top line could be segmented. Morphological or syntactic structure might be so straightforward that a separate line for literal translations, like the fourth line in (1), is unnecessary.

Finally, various additional kinds of information about the example may be useful. In (1), the speaker has been identified. Some speakers are willing to be credited for their contribution to the work, and their identity can provide important information about differences among dialects, ages, genders, etc. Other speakers prefer to remain anonymous, and of course their preferences should be respected. In this example, there is also an indication of where the sentence can be found in the corpus: it occurred 2 minutes and 43 seconds into a narrative now identified as the Bridge text. Such annotation can allow readers to check things for themselves and access further information such as discourse context and prosody.

**3. PHONETICS AND PHONOLOGY.** Examples of words containing each of the distinctive sounds in the language and their variants have long been a basic component of most grammars. Now that audio files can be included with grammars, and grammars can be published in electronic formats with embedded sound, more phonetic information can be included with the description. The accessibility of audio data is a wonderful advance for all readers, both those hoping just to learn about the language and those hoping to learn to speak it. Advances in tools for acoustic analysis are making new kinds of visual displays possible, such as vowel spaces and pitch traces.

The potential value of such displays can again be illustrated with Mohawk. The language shows a two-way tone contrast on stressed syllables. The tones are not simply level high and level low. Each has a distinctive pitch contour or melody. In long, stressed syllables, what is referred to as high or rising tone consists essentially of a rise in pitch. What is referred to as low or falling tone first rises more quickly to a point higher than a basic rising tone, then plunges steeply to a point below the baseline. The effect can be described in words, as here, but, a pitch trace can make things clearer. A comparison of the two pitch contours can be seen in Figure 1 created with Praat software (<http://www.fon.hum.uva.nl/praat/>). The word *onón:ta'* 'hill' [onú:daʔ] with rising tone was pronounced twice, followed by the word *onòn:ta'* 'milk' [onù:daʔ] with 'falling' tone, also pronounced twice.

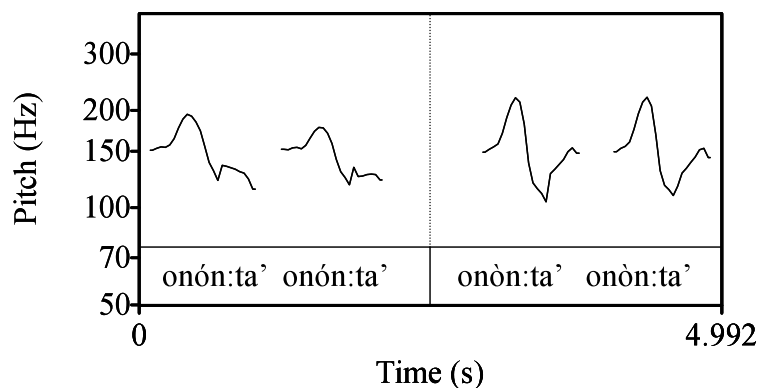


FIGURE 1: Rising versus falling tone contours in Mohawk.

Many languages show special phrase-medial or phrase-final phonological effects. After recording words in isolation, eliciting them in a carrier phrase so that the target word is phrase-medial can be revealing. For interested grammar readers, pitch traces can again provide an effective visual display. In Mohawk, stress is basically penultimate. Open syllables are lengthened and carry one of the two pitch contours seen in Figure 1 above. When a word with penultimate stress and rising tone on an open syllable is followed by another word, however, the pitch continues to rise into the next syllable. This effect can be seen in Figure 2. The first phrase consists of two words: *Sonkwehón:we kenh?* ‘Are you Native?’ The second is a single word, nearly the same as the first: *Konkwehón:we* ‘I am Native’. Both show rising tone and length on the penultimate syllable *hón*. (The orthographic digraph *on* represents a high back nasalized vowel [u̠].) Phrase-medially however, as in the question here, the pitch continues to rise higher into the posttonic syllable *we*: *Sonkwehón:wé kenh?*

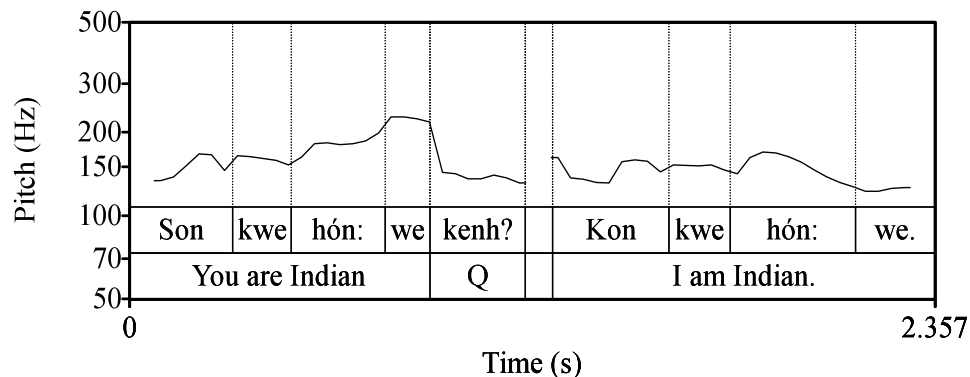


FIGURE 2: Mohawk pitch in context.

**4. MORPHOLOGY AND LEXICALIZATION.** Readers generally expect a good reference grammar to list all affixes or morphological processes, with their ranges of form and function. They expect a complete description of the variant shapes of each. Particularly where a grammar is going to serve multiple purposes and audiences, examples of all of these are important.

The point can again be illustrated with Mohawk. All Mohawk verbs contain a pronominal prefix identifying their core arguments: one argument for intransitives, two for transitives. I may discover that the second person dual prefix ‘you two’ of imperatives has the form *seni-* before *k*, but *tsi-* (IPA [dʒj]) before *a*.

(3) Mohawk dual pronouns

<i>seni-ká:we</i>	‘Paddle, <u>you two</u> !’
<i>tsi-atkátho</i>	‘Look, <u>you two</u> !’
<i>tsi-átien</i>	‘Sit down, <u>you two</u> !’

I should check to see what shape this prefix takes before other consonants and vowels. I may look through my data and discover that the form *seni-* also occurs with all other consonants in the language. Rather than simply stating this fact, it is good to provide examples of each.

## (4) Mohawk dual pronouns before consonants

<i>seni-tákhe</i>	'Run, <u>you two</u> !'
<i>seni-hrárho</i>	'Pull ashore, <u>you two</u> !'
<i>seni-ráthen</i>	'Climb up, <u>you two</u> !'
<i>seni-nóhare</i>	'Wash it, <u>you two</u> !'
<i>seni-niá:ken'n</i>	'Escape, <u>you two</u> !'

But verb stems beginning with other vowels are rarer, and there may be no examples even in an extensive corpus of them with second person dual pronominal prefixes. If I have identified specific stems that begin with the crucial vowels, Mohawk speakers easily provide dual commands with them, so long as the combinations make sense.

## (5) Mohawk dual pronouns before vowels

<i>sen-itskó:tak</i>	'Stay seated, <u>you two</u> !'
<i>sen-è:iahre</i>	'Remember, <u>you two</u> !'
<i>sen-ó'kwat</i>	'Dig, <u>you two</u> !'

If we generalized from the form in (3), we would be wrong—the form *tsi-* only occurs before the vowel *a*.

Restricting the number of examples to a small set can make a grammar quicker to read. In some cases it can also make it more difficult for readers to grasp the principle in question, particularly when the description is couched in a very formal framework. And it can result in missed opportunities: there may be complexities that are not evident from one form alone, complexities that were not anticipated at the time of writing. In addition, language teachers need more than one example to teach a pattern. Especially for purposes of language revitalization, rich exemplification is crucial.

Full inflectional paradigms, in which the stem is kept constant, are typically appreciated and heavily used. Presenting information about forms in two places, once grouped by category (second person dual), and once summarized in paradigms, may not be maximally parsimonious, but particularly where the patterns are complex, it can be space well spent.

There is a traditional expectation that inflection is fully productive, that forms exist for all possible combinations of stems and pertinent inflectional categories in a language. But gaps in inflectional paradigms do exist, and they can be difficult to discover from spontaneous speech alone: the forms in question simply never occur. They can sometimes be uncovered through elicitation, but it is crucial that the elicitation be sensitive. Speakers need to be aware that their sense of what actually exists in the language is precious. Specification of the distinction between possible and actual forms is a valuable part of the grammar. Gaps can be significant.

As noted, all Mohawk verbs contain a pronominal prefix identifying their core arguments, as in *rák-hsere*'s 'he is chasing me'. Similar prefixes can be seen in kinship terms. Relatives are described in Mohawk not as possessions (like the English *my grandfather*), but as relationships. The senior member of a relationship is indicated much like the grammatical agent of a verb, and the junior member much like the grammatical patient of a verb.



Verb	Kinship term
(6) <i>rák-hsere's</i> M.SG/1SG-chase-STATIVE.DISTR ' <u>he</u> is chasing <u>me</u> '	<i>rak-hsót=ha</i> M.SG/1SG-be.grandparent.to=DIMINUTIVE ' <u>he</u> is grandparent to <u>me</u> ' = 'my grandfather'
(7) <i>ri-hsere's</i> 1SG/M.SG-chase-STATIVE.DISTR 'I am chasing <u>him</u> '	<i>ri-ièn'=a</i> 1SG/M.SG-have.as.offspring=DIMINUTIVE 'I have <u>him</u> as offspring' = 'my son'

The prefixes distinguish three persons, three numbers, and four genders. As a result, paradigms can be large: *shonkenihsótha* 'we two, our grandfather' = 'he is grandparent to us two' (M.SG/1DU), *ietshihsótha* 'you all, your grandmother' = 'she is grandparent to you all' (1SG/2PL), *shakotihsótha* 'their grandparents' = 'they are grandparent to them' (M.DP/3DP), and many more.

Gender is distinguished only in third person in Mohawk.

(8) Mohawk genders	
Masculine	Male persons, a few obviously male animals such as 'bull', 'rooster'
Neuter	Inanimate objects
Zoic	Most animals, some female persons
Indefinite	Generic persons ('one, they') and other female persons.

Two different genders are used for referring to female persons: Zoic and Indefinite (also termed Feminine-Zoic and Feminine-Indefinite). The factors underlying the choice between the two are subtle and intriguing. All speakers agree, however, that when discussing one's grandmother or mother, only Indefinite verbs are used, as a sign of respect. For 'she is sewing', one would never say *ka-'níkonhs* (*ka-* 'she ZOIC'), only *ie-'níkhonhs* (*ie-* 'she INDEFINITE) if referring to one's grandmother. For 'I like her', one would never say *ke-nòn:we's* (*ke-* 1SG/ZOIC), but only *khe-nòn:we's* (*khe-* 1SG/INDEFINITE).

Paradigms of kinship terms are very large, as noted, but they show some surprising inflectional gaps. There are no Indefinite kinship terms for grandmothers or mothers, even though the forms they would have if they did occur are obvious.

(9) Kinship term gaps	
no	<i>ak-hsótha</i> 'she (ZOIC) is grandparent to me' = 'my grandmother'
	* <i>ionk-hsótha</i> 'she (INDEFINITE) is grandparent to me'
no	<i>ake-'nisténha</i> 'she (ZOIC) is mother to me' = 'my mother'
	* <i>ionke-'nisténha</i> 'she (INDEFINITE) is mother to me'

These gaps are particularly surprising in light of the fact that in verbs, only Indefinite forms are used for these very persons. In fact the prefixes on kinship terms and associated

verbs do not match. In (10) below, the prefix on ‘my grandmother’ is Zoic, but the prefix on ‘she is sewing’, referring to the same person, is Indefinite.

(10) Kinship mismatches

*Ak-hsótha*

ak-hsot=ha

Z/1SG-be.grandparent.to=DIMINUTIVE

she (Z) is grandparent to me

‘My grandmother is sewing.’

*ie-’níkhonhs.*

ie-’nikhon-hs

I-sew-HAB

she (I) is sewing

It turns out that the verb forms are an innovation. The Indefinite category was originally used only as a generic: ‘one, they’. It retains this function in all of the modern languages in the family. At a certain point, however, people started using it as a sign of deference to certain women. This innovation has now worked its way through the entire Mohawk verbal paradigm: all verbs have both Zoic and Indefinite forms. But it has not fully penetrated the kinship paradigms. The Mohawk terms for ‘my grandmother’ and ‘my mother’ are hold-outs: only the original Zoic forms are used. (These originally referred to all third persons, before the introduction of special Masculine pronouns.) The lag in development of the grandmother and mother forms is not altogether surprising. These are among the earliest terms learned and used by children, and among the most frequent. They are learned as chunks and deeply entrenched, rather than assembled online, thus making them especially resistant to change.

A sensitivity to the difference between actual and possible forms is crucial. Actual forms are those that are known and used spontaneously by speakers. Possible forms are those that a speaker may be able to create on demand. As morphological patterns are uncovered, it can be easy for everyone to be carried away with the magical regularities. Speakers can unwittingly create new forms by analogy, without considering whether the forms are actually used. A grammar should provide a record not simply of what could exist in the language, but of what speakers recognize as established entities in their lexicon, words they have heard. Careful elicitation practices are key here: speakers need to know that their intuitions about what is actually said are valued.

In the domain of word formation (derivation and compounding), the distinction between actual words and possible words is all the more important. Derived forms that actually exist provide a record of concepts that speakers have codified. It should go without saying that non-native grammarians should not invent forms, particularly derived forms, even when they feel they have understood the general principles in play. The verb in (11) was provided by one linguist to demonstrate that ‘unaccusative’ verbs allow incorporation of their ‘subjects’.

(11) Mohawk word?

*Wa’-ka-wí:r-en’-ne’*

FACTUAL-N.SG-baby-fall-PFV

‘The baby fell.’

Surprised to see this example in print, I consulted a group of speakers. Their reactions were strong. “That’s absolutely terrible!”, “Even the worst speaker wouldn’t say this!”, “Oh my gosh!” (and worse). They commented that this word had obviously never been presented to any speaker for approval. Even though Mohawk is polysynthetic, with many long, complex words and highly productive morphology, speakers have a keen sense of which words are part of the language and which are not. The verb stem used when an animate being falls is actually *-ia’t-en-’n-* ‘body-lie-INCHOATIVE’, with incorporated noun stem *ia’t-* ‘body’.

(12) Mohawk word

*Wa’-ka-ia’t-en-’n-e’.*

FACTUAL-N.SG-body-lie-INCHOATIVE-PFV

‘She fell.’

This stem *-ia’t-en-’n-* is lexicalized, a recognized vocabulary item. It reflects a larger pattern whereby verbs are classified as having physical, mental, or abstract effects, by means of the incorporated nouns *ia’t-* ‘body’, *’nikonhr-* ‘mind’, or *-rihw-* ‘matter’. It is especially unfortunate when invented examples like that in (11) become part of the published record of an endangered language. If examples are chosen from a corpus of actually occurring forms, they can serve as a valid resource for those seeking to learn the language, as well as for those seeking to discover further generalizations about the language, like the body/mind/matter classification.

**5. SYNTAX.** It has been observed that descriptive grammars written during the first half of the 20th century tended to concentrate on phonology and morphology, while those written later include more extensive discussions of syntax (Cristofaro 2006:138, Rice 2006:239, and others). Much of this difference can be attributed to progress in our understanding of the kinds of syntactic constructions that exist in different languages and the ways they differ. But when describing and exemplifying syntactic structures, even greater care must be taken in selecting data than when describing phonology and morphology. Phonological and morphological structures tend to be more automated, with components usually below the consciousness of speakers. For the most part, speakers produce complex words on demand without danger of phonological or morphological ‘mistakes’: few stumble over choices among allophones or allomorphs. Strings of words, however, are another matter.

One way in which elicited and constructed examples can be problematic is the distribution of information over lexical categories. A linguist offered the example in (13) as evidence that Mohawk *-’ke* is an adposition. (Spelling, glossing, and analyses have been adjusted here to conform with community conventions and standard analyses.)

(13) Mohawk adposition?

<i>Thí:ken</i>	<i>o-nont-á-’ke</i>	<i>ió-hskats</i>	<i>ne</i>	<i>o-kwir-e’=shòn:’a.</i>
that	N.SG-hill-LK-place	N.SG-be.pretty	the	N-tree-NOUN.SUFFIX=DISTR
‘On that hill, the trees are pretty.’				

The argument being made was that NPs cannot be adjoined to a clause unless they bind some gap or pronoun inside that clause, but that NPs governed by an adpositional element are not subject to this restriction. The sequence ‘on that hill’ must thus be an adpositional phrase. Even if the sentence were acceptable (which speakers agree it is not), it would not prove the principle. The morpheme *’ke* is actually a derivational suffix that derives nouns referring to places. It is not relational: it does not specify a grammatical relation like case endings or adpositions in other languages. Mohawk nouns with such derivational endings can serve as core arguments of clauses, not just adverbs: ‘I know that town (town-place)’, etc. A reasonably-sized corpus of unscripted speech would show this. Some placenames contain this ending, and some do not. Furthermore, derived terms for places can even serve as the input to further derivation. The name of one Mohawk community, for example, is *Kahnawà:ke*. The term for residents of the community was formed by the addition of another derivational ending. The term for ‘heaven’ is literally ‘sky place’. The term for ‘angel’ is literally ‘sky place resident’.

- |                                    |                                  |
|------------------------------------|----------------------------------|
| (14) <i>Kahnawà:ke</i>             | <i>Kahnawa’kehró:non’</i>        |
| ka-hnaw-a-’ke                      | ka-hnaw-a-’ke=hronon’            |
| N-rapids-LK-place                  | N-rapids-LK-place=resident       |
| ‘Rapids place’ = <i>Kahnawà:ke</i> | ‘ <i>Kahnawà:ke</i> resident(s)’ |
| <br>(15) <i>karonhià:ke</i>        | <br><i>raronhia’kehró:non’</i>   |
| ka-ronhi-a-’ke                     | ra-ronhi-a-’ke=hronon’           |
| N-sky-LK-place                     | M.SG-sky-LK-place=resident       |
| ‘heaven’                           | ‘angel’                          |

The creation of the sentence in (13) above seriously distorts the syntactic structure of Mohawk.

Because most good grammars are going to be consulted by an ever-widening range of readers with varying purposes, it is crucial that all data in them be accurate on every level. The sentence in (13) was published with errors in its transcription, segmentation, analysis, and glossing. But perhaps more serious is the fact that even if the individual words were accurate, the sentence is not Mohawk. One speaker explained politely, ‘That’s not something that would ever be said. Perhaps it was written by someone trying to learn. You might forgive them for that.’ When asked what she might say in such a situation, she suggested (16).

- (16) Mohawk counterpart: Kaia’titáhkhe’ Jacobs, speaker p.c.
- |                  |                 |
|------------------|-----------------|
| VERB             | DEM             |
| <i>Ióhskats</i>  | <i>thí:ken,</i> |
| io-hskats        | thiken          |
| N.PATT-be.pretty | that            |
| it is pretty     | that            |

VERB  
*tiokwiró:ton'*.  
 t-io-kwir-ot-on-'  
 CISLOCATIVE-N.PAT-tree-stand-DISTR-STATIVE  
 there it tree stands here and there

This sentence is instantly recognizable as robustly Mohawk. It exemplifies a common pattern of expression. Speakers manage the flow of information such that each significant new idea is introduced in a separate intonation unit or prosodic phrase. In this construction, a basic idea is first introduced by a verb, a complete clause in itself: 'it is pretty', followed by the demonstrative *thí:ken* 'that'. The demonstrative serves as a sort of place-holder, signalling that further elaboration is to come. The distribution of information over words is entirely different from that in (13), which contains a verb and two noun phrases, each with a determiner; (16) consists of two verbs plus a demonstrative. The speaker who suggested (16) noted that she could not imagine a situation where the hill would be mentioned in the same sentence. She hypothesized that if for some reason it were needed, it would be introduced in another sentence.

The linguist who produced the tree sentence in (13) also produced the sentence in (17) below as an example of a predicate nominal construction. (Again spelling and analyses have been regularized.)

(17) Mohawk lexical categories?

<i>Kanónhsa'</i>	<i>thí:ken</i>	<i>o'nerohkwa'kénha'</i> .
ka-nonhs-a'	thiken	o-'nerohkw-a'=kenha'
N-house-NOUN.SUFFIX	that	N-box-NOUN.SUFFIX=DECESSIVE
'That old box is a house.'		

Asked about this one, speakers all agreed that someone might come up with this if they were just learning the language and trying to translate from English. The word *kanónhsa'* has the morphological structure of a Mohawk noun, but this word is not normally used for real houses. Nouns for immovable entities like buildings are incorporated into verbs. The normal way to refer to a house, if it is not incorporated into another verb, is in (18). As one speaker commented, 'A house can't just be there in a vacuum'.

(18) Standard Mohawk term

*kanónhsote'*  
 ka-nonhs-ot-e'  
 N-house-be.standing-STATIVE  
 '(standing) house'

The Mohawk speaker who provided (16) above suggested that to try to convey the meaning apparently intended for (17), one might say something like (19).

(19) Mohawk idiomatic alternative: Kaia'titáhkhe' Jacobs, speaker p.c.

<i>Wahatenonhsónnia'te'</i>	<i>thí:ken,</i>
wa-ha-ate-nonhs-onni-a't-e'	thiken
FACTUAL-M.SG.AGT-MIDDLE-house-make-INSTR.APPL-PFV	that
he house.made himself with it	that
'He made himself a house with it, that thing,	
 <i>o'nerohkwa'kénha'.</i>	
o-'nerohkw-a'=kenha'	
N-box-NOUN.SUFFIX=DECESSIVE	
former box	
box.'	

This is the same Mohawk construction seen earlier, actually more interesting grammatically than that in the invented example in (17). The first prosodic phrase ends in the demonstrative *thí:ken* 'that', promising further elaboration to follow. It shows that ideas expressed in nouns in English are often expressed in verbs in Mohawk. It also shows a different conception of the argument structure, in which humans play the grammatically prominent roles, typical of Mohawk.

It might be countered that (19) would not illustrate the structure that (17) was constructed to show, which is true. But there is a deeper issue here that grammarians should consider. Languages can differ syntactically in ways beyond word order or conditions for omitting pronouns. They can vary in how ideas are distributed over lexical categories, over predicates and arguments, over clauses, and over sentences. If the data in a grammar consist only of constructed examples and sentences translated from another language, we stand to miss much of what that language has to teach us beyond what we already know. In her article 'Writing culture in grammar in the Americanist tradition', Jane Hill (2006) eloquently discusses various ways grammars might reflect the culture of which a language constitutes a part. We can do more than provide translation equivalents of what we consider theoretically relevant sentences in another language. We can strive to capture glimpses of how speakers package their thoughts, what they choose to say and how they choose to say it.

The choice of examples of complex sentences raises similar issues. A substantial corpus, with a variety of kinds of speech represented, is likely to provide ample exemplification of complex constructions, probably of more types and greater complexity than one might think to elicit. At the same time, not all constructions one might think of will necessarily surface. Another linguist recently noted the theoretical importance of Chomsky's 1977 article 'On *wh*-movement', suggesting that investigating parallels in other languages might lead to further understanding of their syntax. Chomsky contrasted the two English sentences in (20) to show that movement of the *wh*-word to the front of a sentence is unbounded, as in (20a), but that the presence of a noun phrase like *the story* in (20b) blocks the movement because of a subadjacency violation. The brackets [ ] indicate the position from which the question word is said to have moved.

- (20) a. What did Susan say Mary thought John should persuade Bill to buy [ ]?  
 b. \**What* do you believe the story that Mary bought [ ]?

For some languages, translating these sentences then asking for grammaticality judgments might produce clear answers and interesting results. For others, it risks clouding the picture. Just as progress has been made in our understanding of syntactic patterns, so too, are advances being made in our understanding of discourse patterns. It is not surprising that sentences comparable to those in (20) do not occur in even a substantial corpus of unscripted Mohawk speech. Speakers do not combine long strings of ideas like those in a single sentence, for systematic reasons involving information structure. Should we fill in the gap with elicitation? It might be possible to find a Mohawk speaker who could be persuaded to translate them into Mohawk, or to give grammaticality judgments about an interviewer's translation of them. But we might ask what such judgments would represent. Speakers of most endangered languages are bilingual: that is usually part of the story of endangerment. There are still excellent, articulate Mohawk speakers, though all of them are also good speakers of English. Faced with hypothetical sentences like these, even the speakers themselves cannot know how much they are tapping into their intuitions about English. If Mohawk translations of such sentences became part of the record of the language, they would certainly misrepresent its discourse and information structure, that is, the way speakers actually distribute information over clauses and sentences.

**6. DISCOURSE AND INTERACTION.** Spontaneous connected speech offers insight into fundamental features of a language, in many cases the kinds of features that make the language special. All connected speech is not the same, however. Particular features often show different degrees of elaboration in different kinds of discourse. It is important to draw examples for a grammar not just from monologue, but also from interactive conversation, normally the kind of speech that is both the most frequent and of most relevance to revitalization projects.

The importance of genre can again be illustrated with examples from Mohawk. The passage in (21) opened a story written by a group of Mohawk language teachers, all excellent first-language speakers.

(21) VERB

*Tewakhwishenhé:ion*

te-wak-hwish-enhei-on

DUPLICATIVE-1SG.PAT-strength-die-STATIVE

my strength has died

'I was tired

PARTICLE

*sok*

sok

so.then

so then

so I quickly went to bed.

VERB

*iohsnó:re'*

io-hsnor-e'

N-be.fast-STATIVE

it is fast

VERB

*onkità:wha'.*

wa'-w-ita'w-ha-'

FACTUAL.1SG.PAT-sleep-ANDATIVE-PFV

I went to sleep

PARTICLE	VERB
<i>Sok</i>	<i>wa'-k-atà:sw-aht-e'</i>
sok	wa'-k-ata'sw-a-ht-e'
so.then	FACTUAL-1SG.AGT-go.out-LK-CAUS-PFV
so then	I extinguished
Then I turned off my light	

PARTICLE	VERB
<i>tanon'</i>	<i>ia'ká:rate'.</i>
tanon'	ia'-k-arat-e'
and	TRANSLOCATIVE-1SG.AGT-lie-PFV
and	I lay down there
and lay down.'	

Morphologically and syntactically, the passage is fine. The words are all well formed and idiomatic. The distribution of information over nouns and verbs is typical of Mohawk speech; there are few nouns in the entire story. There are just two particles that relate ideas to each other: *sok* 'so then' and *tanon'* 'and'.

Compare the style of (21) above to that of (22) below, from a conversation. A group of friends were discussing an old man they had known as children.

(22) Charlotte Bush, speaker p.c., Onkw A 41.00

PARTICLE	VERB
<i>Tanon'</i>	<i>raonkwe'táksen.</i>
tanon'	raw-onkwe-'t-aks-en
and	M.SG.PAT-be.a.person-NMLZ-be.bad-STATIVE
and	he is person bad
'And he was cross.'	

Watshenní:ne' Sawyer, speaker p.c.

VERB	PARTICLE	PARTICLE
<i>Rorihwakwénienhs</i>	<i>nen'</i>	<i>nè:'e;</i>
ro-rihw-a-kweni-enhs	nen'	nè:'e
M.SG.PAT-matter-LK-be.able-HAB	that	that
he is matter competent	CONTRASTIVE	
'He was respectful though;		

VERB	PARTICLE
<i>rorihwakwenienhstòn:ne</i>	<i>nek tsi</i>
ro-rihw-a-kweni-enhst-onhne	ne-k tsi
M.SG.PAT-matter-LK-be.able-HAB-PAST	the-only as
he had been matter competent	but
he used to be respectful but	



PARTICLE	PARTICLE	PARTICLE	VERB	PARTICLE
<i>khere'</i>	<i>kati'</i>	<i>kenh</i>	<i>thitewana'kón:nihsuwe'</i>	<i>wáhi'</i>
khere'	kati'	kenh	thi-te-wa-na'kw-onni-hskwe'	wahi'
perhaps	in.fact	Q	CONTR-1INCL.AGT-PL-anger-make-PAST.HAB	TAG
I guess	in fact	Q	we used to make him mad	TAG
I guess in fact we used to make him mad, didn't we.'				

This passage contains noticeably more particles than that in (21). In the first line uttered by the second speaker, the sequence *nen'nè:'e* marks a focus of contrast with the preceding comment. In the second line, the particle *neksi* 'but' announces a contrast to come in the third line. In the third line, the particle *khere'* is a sort of inferential indicating that the speaker is imagining the reason behind the man's crossness. The particle *kati'* signals that this sentence is relevant to the preceding discussion. The particle *kenh*, normally the yes/no question marker, here serves to suggest slight doubt. The final particle *wáhi'* is a tag, comparable to English 'isn't it' or here 'didn't we'. This is an example of one of the common interactive uses of the Mohawk tag *wahi'*. This was a co-constructed narrative. The tag signaled that the speaker was not setting herself up as the only one knowledgeable about the topic, and solicited participation from the others. She was successful: her comment was met with affirmation from the others, one of whom then added to the story.

Data from elicited translations, and even from texts constructed by speakers sentence by sentence as in (21) above, are typically poor in the devices that speakers use spontaneously to shape their messages, highlight significant information, background peripheral or familiar ideas, mark contrasts with previous statements or general knowledge, show links to previous threads of discussion or fresh starts, interact with their audience, and much more.

**7. PROSODY BEYOND THE WORD.** A key element of linguistic structure is prosody. Technological advances have made it possible to observe patterns of pitch, intensity, rhythm, and phonation closely and even quantitatively if desired, and to include descriptions of these patterns in grammars. Examples were seen in Section 3 of visual displays of the pitch countours associated with distinctive tones on Mohawk words. Similar displays can add helpful information to descriptions of larger stretches of speech. The relation between grammar and prosody is not isomorphic: in some cases the two run in parallel and reinforce each other, but in others they convey different information. Neither can be fully predicted from the other.

Prosody is typically a significant component of question constructions, for example, but the prosody of questions varies across languages. With acoustic analysis, we can see the prosodic patterns that accompany different kinds of questions and include visual representations of them in the grammar. English yes/no questions tend to show a rise in pitch. Their Mohawk counterparts generally do not. The difference can be seen by comparing the two frames in Figure 3, produced by the same bilingual speaker. The first shows the pitch contour of the Mohawk question *Tentéhse'kenh?*. The peak occurs on the stressed syllable of the first word. The contour contrasts sharply with its English counterpart 'Will you be back?', which shows a steady rise.

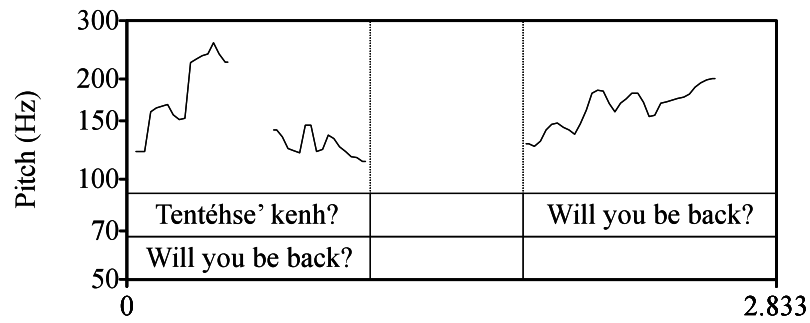


FIGURE 3: Mohawk and English prosody

Mohawk question-word questions show a similar fall in pitch, as can be seen in Figure 4.

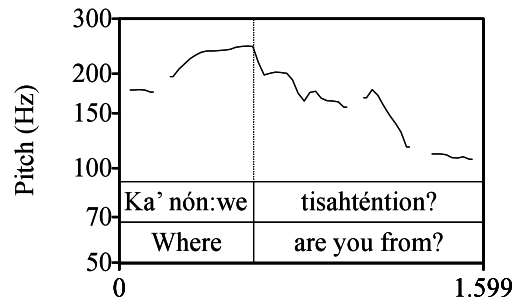


FIGURE 4: Mohawk lexical gap question

Tag constructions with *wáhi'* also show a final fall in pitch, as can be seen in Figure 5, a pitch trace of the final line of example (22) above.

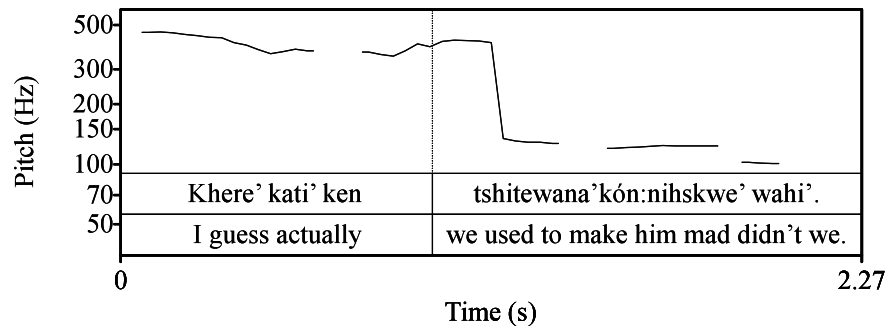


FIGURE 5: Falling pitch in tag construction

Prosodic patterns can also be revealing for studies of syntactic complexity. Looking at words in print, one might come to the conclusion that Mohawk speakers simply speak in sequences of simple sentences. Consider example (23) below.

(23) Mohawk: Billy Two Rivers, speaker p.c., B2R 50.17

<i>Wahskwé:ni'</i>	<i>á:re'</i>	<i>thé:nen'</i>	<i>wahsì:ron'</i>	<i>kenh?</i>
wa-hs-kweni-	are'	othenen'	wa-hs-ihron-	kenh
FACTUAL-2SG.AGT-be.able-PFV	again	something	FACTUAL-2SG.AGT-SAY-PFV	Q
you managed	again	anything	you said	Q
'You couldn't manage to say anything different, eh?'				

Both of the verbs, *wahskwé:ni'* 'you managed' and *wahsì:ron'* 'you said it', are finite and could be used on their own as independent sentences.

*Wahskwé:ni'*.  
'You managed it.'

*Wahsì:ron'*.  
'You said it.'

The free translation, later provided by another speaker who participated in the conversation, indicates that the utterance was understood as one complex sentence. The prosody shows the same structure: the two clauses 'you managed' and 'you said something' were integrated under a single intonation contour, with a pitch reset only at the beginning, on the stressed syllable of the first word of the first clause (*wahskwé:ni'* 'you managed'), and a continuous fall until the end of the second clause. (The precipitous drop here is due to the falling tone on *wahsì:ron'* 'you said it'.)

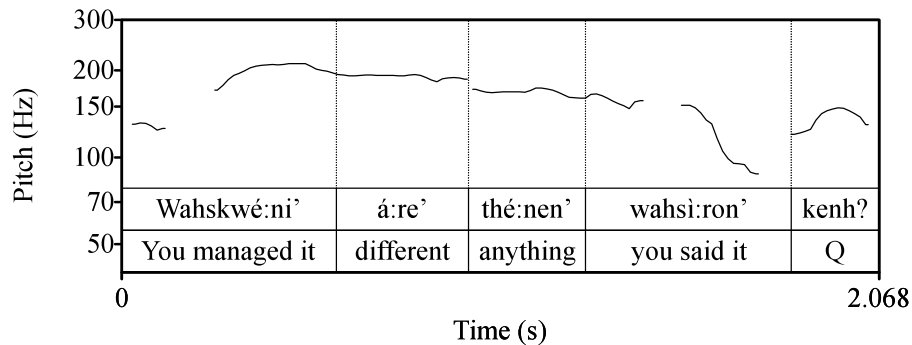


FIGURE 6: Single integrated prosodic envelope for complex sentence

Examples of prosodic structure are all the more important in descriptions of discourse and information structure. The basic unmarked prosody of a Mohawk sentence shows a

progressive descent in pitch from one stressed syllable to the next. The pitch trace for example (24) is in Figure 7.

(24) Basic simple sentence: Watshenní:ne' Sawyer, speaker p.c., Onkw A 3.35

<i>Ó:nenhste'</i>	<i>ken'k</i>	<i>nikontihnenié:son's</i>	<i>tanon'</i>
o-nenhst-e'	ken'=k	ni-konti-hneni-es-on's	tanon'
N-corn-NOUN.SUFFIX	small=only	PARTITIVE-Z.PL-height-be.long-STATIVE.DISTR	and
corn	just small	so they are variously tall	and
'The corn is very short and [it all seems to be doing poorly].'			

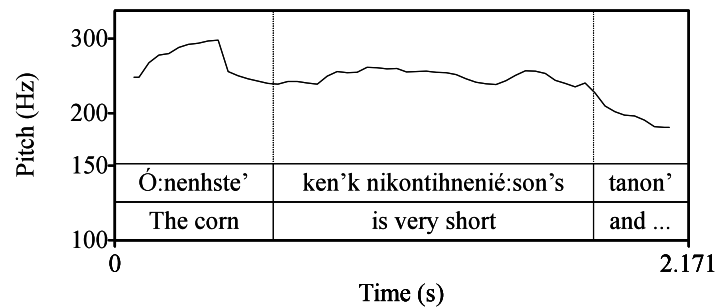


FIGURE 7: Basic declination over a prosodic phrase

When examined in print alone, the example in (25) appears to show the same structure.

(25) Topicalization: Watshenní:ne' Sawyer, speaker p.c., Onkw A 41.15

<i>Akhsotkénha'</i>	<i>wahonwatinónhsani'</i>	<i>iatathróna'</i>
ak-hsot=kenha'	wa-honwati-nonhs-a-ni'	i-atat-hrona'
Z/1SG-be.gp.to=DECESSIVE	FACTUAL-I/3PL-house-LK-lend-STATIVE	M.DU-REFL-be.with-STATIVE
my late grandmother	she house lent them	they two are with each other
'My late grandmother rented a house to a couple.'		

But this is a topicalization construction, used when the discourse topic is shifted to a different referent. The construction is usually used when the new topic has already been mentioned or is associated with another referent under discussion. This sentence was part of a lively conversation among a group of half a dozen people. It was the opening to an anecdote. The speaker had just said "I have another story". She then continued to recount her grandmother's experience.

This topicalization construction shows a distinctive pitch contour. It begins on a high pitch, but after the topicalized element, here *Akhsotkénha'* 'my late grandmother', there is a brief break, then a pitch reset on the stressed syllable of the following clause, here *nón*.

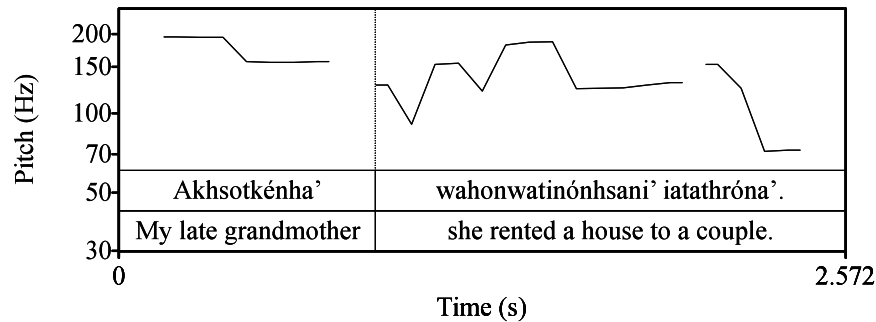


FIGURE 8: Topicalization

The previous example in (24), 'The corn is very short', was not a topicalization construction. The speakers had been discussing a trip they had taken the day before and the things they had noticed along the way. The corn was not mentioned again.

A number of other constructions are distinguished essentially by intonation. Now that the tools are available for us to see the patterns and display them, it makes sense to consider including such displays in grammars.

**8. TYPOLOGY AND THE SELECTION OF EXAMPLES.** An awareness of the kinds of grammatical categories and patterns that recur crosslinguistically is an important tool for grammar writers, making it possible for them to recognize patterns more quickly, identify points of general linguistic interest, and know which issues to pursue further. It is also useful when it comes to selecting examples for the grammar. In her 2006 article 'The organization of reference grammars: A typologist user's point of view', Sonia Cristofaro provides a good illustration of such benefits. As she notes, Givón (1980, 2001) and others have observed that the forms of complement constructions can vary across and within languages.

(26) Complementation scale for English: Givón (2001:43)

- |      |                           |  |
|------|---------------------------|--|
| i.   | Co-lexicalized complement | <i>She <u>let go</u> of the knife.</i>                   |
| ii.  | Bare-stem complement      | <i>She <u>let</u> him <u>go</u> home.</i>                |
| iii. | Infinitive complement     | <i>She <u>wanted</u> him <u>to leave</u>.</i>            |
| iv.  | For-to complement         | <i>She'd <u>like for</u> him <u>to leave</u>.</i>        |
| v.   | Subjunctive complement    | <i>She <u>wished</u> that he <u>would leave</u>.</i>     |
| vi.  | Indirect quote complement | <i>She <u>said</u> that he <u>might leave</u> later.</i> |
| vi.  | Direct quote complement   | <i>She <u>said</u>, "He <u>might leave</u> later."</i>   |

When multiple complement constructions coexist within a language, various factors affect their distribution. One is the semantics of the matrix predicate.

Noonan (1985, 2007) distinguishes the following classes of complement-taking predicates.

## (27) Complement-taking predicate types: Noonan (1985, 2007)

- i. Utterance: *say, tell, report, promise, ask*
- ii. Propositional attitude: *believe, think, suppose, assume, doubt, deny*
- iii. Pretence: *imagine, fool, pretend, make believe*
- iv. Commentative (factives): *regret, be sad, be odd, be significant, be important*
- v. Knowledge: *know, discover, realize, find out, forget, see, hear, dream*
- vi. Fear: *be afraid, fear, worry, be anxious*
- vii. Desiderative: *want, wish, desire, hope*
- viii. Manipulative: *cause, force, make, persuade, tell, threaten, let, order, request, ask*
- ix. Modal: *can, be able, ought, should, may, be obliged*
- x. Achievement: *manage, dare, remember to, happen to, get to, try, forget to, fail, avoid*
- xi. Phasal (aspectual): *begin, start, continue, keep on, finish, stop, cease*
- xii. Immediate perception: *see, hear, watch, feel, imagine*

Armed with a framework like this, the grammar writer can select examples of complementation such that each matrix type is represented. Examples can first be sought in unscripted speech, then gaps can be checked with judicious elicitation.

**9. LANGUAGE CHANGE AND THE SELECTION OF EXAMPLES.** Languages are constantly evolving, as speakers work to make sense out of the patterns they perceive, repair apparent irregularities, and extend those patterns to express new ideas. Many such developments are gradual. A particular construction might appear first only in certain lexical contexts, then spread to others lexical item by lexical item. A frequently-used collocation may become ever more frequent, used in more and more situations, and as a result become more general in meaning. Erstwhile syntactic constructions may be used so often that they begin to blend and erode phonologically, ultimately evolving into grammatical and discourse markers. This dynamism is a fundamental feature of any living language, one that should be included as part of the description where possible. A rich body of examples from spontaneous speech in a grammar can often provide a snapshot of such processes in progress.

An example of such a process can be seen in developments of the Mohawk verb root *-ehr-* ‘think, believe, want’. Like other verb roots, it appears in verbs in various aspects and tenses, and with all persons and numbers. It often occurs as the matrix verb of complement constructions.

(28) Mohawk verb *-ehr-* ‘think, believe, want’: Joe Deer, speaker, Sose 1.20

<i>Wà:kehré’</i>	<i>akwé:kon</i>	<i>tenkhenonhwará:ton’.</i>
<i>wa’-k-ehr-e’</i>	<i>akwek-on</i>	<i>t-en-khe-nonhwaraton-’</i>
FACTUAL-1SG.AGT- <u>think</u> -PFV	be.all-STATIVE	DUPLICATIVE-FUT-1SG/3PL-greet-PFV
I thought	all	I will greet them
‘I <u>thought</u> I would greet everybody.’		

(29) Mohawk verb *-ehr-* ‘think, believe, want’: Billy Two Rivers, speakerp.c., B2R 59.55

<i>Áhsehrék</i>
a:-hs- <u>ehr</u> -ek
OPTATIVE-2SG.AGT- <u>think</u> -CONTINUATIVE
‘You would <u>think</u>

<i>tóka</i>	<i>ráonha</i>	<i>iahawennà:reke’</i> .
toka	raw-onha	i-a:-ha-wenn-a-hrek-e’
maybe	M.SG.PAT-alone	TRANSLOCATIVE-OPTATIVE-M.SG.AGT-word-LK-push-PFV
maybe	himself	he would word push
maybe he himself would be pushing for the language.’		

(30) Mohawk verb *-ehr-* ‘think, believe, want’, Kaia’titáhkhe’ Jacobs, speaker p.c. B2R 11.54

<i>Kwah</i>	<i>í:kehre’</i>	<i>thi,</i>	<i>á:re’s</i>	<i>kawennókeri.</i>
kwah	i-k- <u>ehr</u> -e’	thiken	are’-s	ka-wenn-okeri
just	PROTHETIC-1SG.AGT-think-STATIVE	that	again-DISTR	N-word-be.gathered
really	I think	that	again	it is word shrunk
‘I really <u>think</u> that the words are compacted.’				

The examples above are from conversation, but similar examples are easy to elicit. In spontaneous speech, however, the verb *í:kehre’* ‘I think’ appears in constructions that differ to varying degrees from prototypical complement constructions. Sometimes what would seem like a part of the complement clause precedes the matrix.

(31) Kaia’titáhkhe’ Jacobs, speaker p.c., B2R 51.10

<i>Í:se’</i>	<i>í:kehre’</i>	<i>sahtentiòn:ne’</i>
ise’	i-k-ehr-e’	sa-ahtenti-onhne’
2	PROTHETIC-1SG.AGT-think-STATIVE	2SG.PAT-leave-STATIVE.PAST
you	I think	you had left
‘You, <u>I think</u> you were away.’		

There is evidence that this verb is taking on modal meaning, indicating less than total certainty on the part of the speaker. The exchange below is interesting for two reasons. One is that the speaker utters the word *í:kehre’* twice in one sentence, once inside of a simple clause. The other is that another participant in the conversation then expressed agreement, but it was not with the apparent matrix ‘I think’, but rather with the apparent complement: ‘they’ve planted pole beans’.

(32) Interaction: Sose 3.30

A	<i>Tanon’</i>	<i>í:kéhre’</i>	
	tanon’	i-k-ehr-e’	
	and	PROTHETIC-1SG.AGT-think-STATIVE	
	and	I think	
	‘And <u>I think</u>		
	<i>ienakarótha’</i>	<i>í:kéhre’</i>	<i>rotiiénthon.</i>
	ie-nakar-ot-ha’	i-k-ehr-e’	roti-ient-hw-on
	I.AGT-pole-stand-HAB	PROTHETIC-1SG.AGT-think-STATIVE	M.PL.PAT-lie-CAUS-STATIVE
	one pole stands	I think	they have planted
	they’ve planted pole beans.’		

B	Én:,	rotiiénthon	kwah	í:ken.
	en:	roti-ient-hw-on	kwah	i-ka-i
	yes	M.PL.PAT-lie-CAUS-STATIVE	just	PROTHETIC-N.AGT-be
	yes	they have planted	really	it is
	Yes, they really have.' (Not: 'Yes, you really do.')			

Further examination of spontaneous speech shows additional developments. The verb now also appears in a much reduced form *khere* with the meaning 'maybe, perhaps'. Speakers no longer feel that it contains the first person pronominal prefix *k-* 'I'. It often occurs as part of the sequence *khere*'*kati*'*ken*, apparently a combination of *khere* 'perhaps' + *kati* 'in fact, actually', and the interrogative *kenh* which adds uncertainty. Speakers are unsure about whether this sequence consists of three elements or just one.

(33) Modality: Joe Dove, speaker: Sose 9.20

<i>Khere</i> ' <i>kati</i> ' <i>kenh</i>			<i>tóka</i> '	<i>sahontenhni:non</i> '.
khere'	kati'	kenh	toka'	sa-hon-aten-hninon-
perhaps	actually	Q	maybe	REPETITIVE.FACTUAL-M.PL.AGT-MIDDLE-buy-PFV
maybe			perhaps	they sold it again
'I guess maybe they sold it.'				

Such progressions of grammatical development are generally not as evident in elicited or translated material. Elicited examples might be simpler and illustrate a basic grammatical point more clearly, but if all examples are elicited or constructed, the dynamism inherent in the language will be missed.

**10. LANGUAGE CONTACT.** The potentially powerful role of contact in shaping grammar is becoming ever clearer, as more detailed documentation of more languages is becoming available. Grammar writers often make a conscious effort to exclude all non-native features from their descriptions and examples. Particularly in the case of endangered languages, it is important to many communities to have a record of the traditional form of the language as it was spoken before the encroachment of a competing majority language. At the same time, bilingualism has long been the norm in many communities, even before contact with larger world languages. It can enrich languages, as bilingual speakers exploit the distinctions offered by two systems in order to express themselves more eloquently in each. But it can also erode the distinctiveness of a minority language, as patterns are remodeled to mirror those of the majority language. In the end it is communities who have the most to say about what they hope to see documented in a grammar: the most traditional patterns of the heritage language, or the modern usage of skilled bilingual speakers.

The difference is not always obvious. Lexical borrowing can be evident, particularly when the source language is well known. Structural borrowing can be more difficult to spot and evaluate, but it can have more profound effects. Bilingual speakers may create patterns in one of their languages modeled on those in the other with no transfer of lexical substance. They may simply increase the frequency of an existing minor pattern in one language to match the frequency of a comparable major pattern in the other, or extend it to more contexts. But if most or all examples in a grammar are obtained through elicitation



and/or translation, the description may not even reflect the current state of the language. It can be difficult to determine whether the similarities are actually representative of the language or an artefact of the methodology.

Possible contact effects can be obscured even when speakers themselves are constructing examples. Even good speakers can produce structures during a translation process they would never utter spontaneously. When one skilled Mohawk speaker assembled a pedagogical grammar, he produced the question and answer pair in (34).

- (34) a.      *Í:seks    kenh    ne    kanà:taro?*  
               you eat    Q        the        bread  
               'Do you eat bread?'
- b.      *Í:keks    tiótkon    ne    kanà:taro.*  
                       I eat        always    the        bread  
                       'I always eat bread.'

The words are phonologically and morphologically well-formed. The question 'Do you eat bread?' correctly illustrates the position of the interrogative particle *kenh*, immediately after the first constituent. But the speaker who created this example would not talk like this.

The question in (34a) shows a word order similar to that in English, where direct objects routinely follow the verb. But constituent order in Mohawk is not governed by syntactic relations. It is pragmatic: essentially, the most newsworthy information appears early in the clause (often after various orienting and other discourse particles). In yes/no questions, the focus of the question appears initially, followed by the interrogative particle *kenh*, as here. But otherwise the word order in the question above is unusual, with its focus on the eating rather than the bread: 'Do you eat the bread?'. The use of the particle *ne* introduces a second complication. Mohawk *ne* often appears in the same kinds of contexts as English *the*, but it has a subtly different function: it indicates that the referent has already been mentioned or evoked in the current discussion: 'the aforementioned'. The question in (34a), presented in the grammar with no previous context, is thus pragmatically odd in several ways. It might be appropriate if you had been telling me that you bake a lot of bread, and I wanted to ask you whether you actually eat that bread. To ask a more general question about whether someone eats bread, a usual form would be one like that in (35), with the bread first and no *ne*.

- (35) *Kanà:taro        kenh    Í:seks?*  
       bread            Q        you eat  
       'Do you eat bread?'

The answer in (34b) above, *Í:keks tiótkon ne kanà:taro* is also unidiomatic, perhaps again reflecting English patterns. The word *tiótkon* 'always' normally supplies important information and tends to occur at or near the beginning of the clause in Mohawk, as in (36).

- (36) *Hén:, tiótkon ne kanà:taro í:keks.*  
 yes always the aforementioned bread I eat  
 ‘Yes, I always eat bread.’

In this case, it is the process of assembling examples that has produced contact effects.

**11. CONCLUSION.** As our technology and understanding of language progress, so too can our ideas about the kinds of data that might be useful to a wider range of users and that can lead to new discoveries. We are learning more about the intimate relations between structure and substance: speakers do not simply know abstract grammatical patterns on the one hand, and lists of morphemes and words on the other. The strengths of bonds between constructions and particular lexical items probably fall along a continuum. We are learning more about relations between structure and context: speakers select morphological and syntactic constructions for a variety of reasons, often involving the discourse context and the interpersonal situation and goals. We are becoming increasingly aware of the role of prosody in syntax and discourse. We are also becoming more conscious of the constantly evolving nature of linguistic structures and the forces that shape them, both language-internal mechanisms and language contact. If the examples in the grammar are accurate on all levels of structure, they should be useful for learning more about all of these areas of inquiry.

There are now many good grammars that provide models of how to choose effective examples. Among the basic guidelines that have been discussed here are the following.

- 1) Nature of the data
  - a. Drawn as much as possible from spontaneous connected speech, in a variety of genres, critically including ample conversation
  - b. Augmented by elicited examples for clear pronunciations of individual words, completeness of descriptions of allomorphy and paradigms, and illustration of contrasting structures
  - c. Representative of the range of known typologically significant variables
  - d. Accompanied by surrounding context where pertinent
  - e. Generous in quantity
  - f. Culturally appropriate, all else being equal
- 2) Presentation of examples
  - a. Interlinear analysis and glossing for languages where this is not immediately obvious. Different amounts of interlinear information are appropriate for different languages. Interlinear lines may show such things as morphological segmentation, underlying forms, morpheme glossing, and/or literal word-by-word translations.
  - b. Where appropriate, references to locations of the examples in texts and/or audio recordings that would allow the reader to see them in their discourse context or hear them.
  - c. Insofar as possible, use of standard glossing conventions such as the Leipzig Glossing Rules.

- 3) Additional aids
  - a. Figures showing such acoustic information as vowel spaces, pitch contours, vowel spaces, etc.
  - b. Traditional paradigm tables

The special value of data from unscripted connected speech was recently brought home to me when I consulted two grammars of the same language. The first provides clear lists of phonemes and allophones, case endings, demonstratives, pronouns, and quantifiers. It contains examples of complex noun phrases combining all of those elements. There are lists of tense, aspect, and mood endings, reflexives and reciprocals, causatives, negatives, and passives. There are examples of simple sentences, conjoined sentences, relative clauses, adverbial clauses, and complement clauses. All of the kinds of structures a typologically-informed grammarian of this period would seek out are exemplified. The second grammar lists the same elements, though transcriptions are not always as clean. But the two grammars differ in a striking way. The first could be a description of a language almost anywhere in the world. The second is immediately obviously a grammar of a language indigenous to California. This second grammar contains numerous affixes and clitics not mentioned in the first, markers even a typologically sophisticated grammarian might not think to look for. They represent categories that are highly developed in languages indigenous to California, languages that are genetically unrelated but that have been spoken by peoples in close contact over centuries. The areal affiliation of the language comes across robustly in the examples, and not just because of mentions of acorns rather than rice. It is obvious from the ideas speakers chose to express, the distinctions they chose to specify, and the distribution of information over words, clauses, sentences, and larger stretches of speech. As might be suspected, examples in the first grammar were drawn almost entirely from elicitation, direct translations of English models, which they generally parallel word for word. Those in the second came from a vast collection of texts of varied kinds. In those it was the speakers who chose what topics to discuss and what to say about them. The writer of this second grammar was acutely tuned into the genetic and areal context of the language, and he was able to note significant similarities and differences between it and its closest genetic relatives and neighbors. What is perhaps surprising is the fact that this second grammar was written nearly a century before the first. Because of its grounding in extensive unplanned speech, it continues to provide valuable information about issues even this alert grammarian could never have thought of at the time.

#### REFERENCES

- Austin, Peter K. 2003 on. *Language description and documentation*. Endangered Languages Project. London, UK: School of Oriental and African Studies.
- Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*. 152-165. Cambridge, UK: Cambridge University Press.
- Chelliah, Shobhana L. and Willem J. de Reuse 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.

- Chomsky, Noam. 1977. On WH-movement. In Peter W. Cullicover, Thomas Wasow, and Adrian Akmajian (eds.), *Formal syntax*, 71-132. New York: Academic Press.
- Cristofaro, Sonia. 2006. The organization of reference grammars: A typologist user's point of view. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 137-170. Berlin: Mouton de Gruyter.
- Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.) 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4:333-77.
- Givón, Talmy. 2001. Verbal Complements and Clause Union. Chapter 12 of *Syntax*. II:39-90. Amsterdam: John Benjamins.
- Grenoble, Lenore & Louanna Furbie (eds.) 2010. *Language documentation: Practice and values*. Amsterdam: John Benjamins.
- Hill, Jane H. 2006. Writing culture in grammar in the Americanist tradition. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 609-628. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 1-30. Berlin: Mouton de Gruyter.
- Mithun, Marianne 2001. Who shapes the record: The speaker and the linguist. . In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*. 34-54. Cambridge, UK: Cambridge University Press.
- Mithun, Marianne 2007. Grammars and the community. In Thomas E. Payne & David J. Weber (eds.), *Perspectives on grammar writing*. 45-70. Amsterdam: John Benjamins.
- Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description II: Complex constructions*. 42-140. Cambridge: Cambridge University Press.
- Noonan, Michael. 2007. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description II: Complex constructions*, 2nd edn. 52-150. Cambridge: Cambridge University Press.
- Rice, Keren. 2006. Let the language tell its story? Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 235-269. Berlin: Mouton de Gruyter.
- Weber, David J. 2007. The linguistic example. In Thomas E. Payne & David J. Weber (eds.), *Perspectives on grammar writing*. 199-213. Amsterdam: John Benjamins.

*mithun@linguistics.ucsb.edu*